subsequent data collection efforts are due. Programs and values may change. Issues initially considered to be of great importance to study stakeholders may fall into disfavor because of changes in the social and political climate. Subjects on whom baseline data were collected may no longer have the same characteristics that caused them to be chosen initially for the study, or they may have done something as obvious as move away from the location of the study.

Theoretical vs. Data Definition of Subjects

Let us consider the definition of study subjects. The study designer identifies subjects in terms of a set of characteristics based on the designer's current knowledge of the research environment and the rationale for the study. Identification would include number and kind in terms of their physical and geographic location. This process is documented in a data collection sample.

It is the job of the data collection staff to identify subjects as particular individuals. This is usually done by a rostering or listing of subjects including unique identification data. Using the information collected in the field, the data processing staff identifies each subject by a coding or ordering system. Finally, the analysis staff may use part or all of the identifiers established by the design, field, and data processing staff to create an infinite number of new subject identifiers.

Attrition may affect each major functional area.

Attrition

Design attrition occurs when subjects initially involved in the study are excluded because of changes in study focus or rationale. Data collection attrition occurs when subjects included in the design are excluded from the data gathering process because they were not contacted by the data gatherer. Data collection attrition also includes subjects who were excluded because the contact with the data gatherer did not yield information required for analysis. Data processing attrition occurs

Methods for Minimizing Attrition in Field Studies

by

Gloria Marshall

Stanford Research Institute

Presented at the Annual Meeting of the American Educational Research Association at San Francisco, 22 April 1976.

Plans vs. Implementation--The Basic Research Problem

Taking a research study from the planning to the operational phase is a delicate task indeed. Data specifications may not take account of real world conditions. Or, realistic data specifications my be incorrectly implemented by the data handling staff. For the sake of study validity, all inaccuracies should be defined and documented. In some instances, the recognition of inadequacies in theory or data handling must be resolved by altering the study plans.

Both theoretical and data handling personnel must remain in touch with changes in the real world that may affect their plans. They must also be ready to document innacuracies and redefine specifications when necessary. And one problem that they must be prepared to deal with in any study that requires more than one collection of data is attrition.

Attrition is the loss of study subjects. It is of several kinds, but it has the same effect regardless of kind: it confounds the subject identification process.

Relationship of Study Characteristics to Resolution of Data Specification Problems

The size of a study and the length of time over which it takes place may affect the resolution of data specification problems. In very small projects, the designer, data collector, and analyst are often the same person or a single small group. Differences between theoretical definitions and implementation of theories, once they have been identified, can easily be worked out by adjusting the design and/or data handling procedures. However, in projects large enough to warrant the division of research tasks among many persons, communication among functional groups becomes more difficult; changes in procedures and alteration of analyses plans become complex organizational problems.

In longitudinal studies, the probability that planning and implementation will be dissonant is increased. Definitions and descriptions set up at the time of baseline data collection may have altered by the time

subsequent data collection efforts are due. Programs and values may change. Issues initially considered to be of great importance to study stakeholders may fall into disfavor because of changes in the social and political climate. Subjects on whom baseline data were collected may no longer have the same characteristics that caused them to be chosen initially for the study, or they may have done something as obvious as move away from the location of the study.

Theoretical vs. Data Definition of Subjects

Let us consider the definition of study subjects. The study designer identifies subjects in terms of a set of characteristics based on the designer's current knowledge of the research environment and the rationale for the study. Identification would include number and kind in terms of their physical and geographic location. This process is documented in a data collection sample.

It is the job of the data collection staff to identify subjects as particular individuals. This is usually done by a rostering or listing of subjects including unique identification data. Using the information collected in the field, the data processing staff identifies each subject by a coding or ordering system. Finally, the analysis staff may use part or all of the identifiers established by the design, field, and data processing staff to create an infinite number of new subject identifiers.

Attrition may affect each major functional area.

Attrition

Design attrition occurs when subjects initially involved in the study are excluded because of changes in study focus or rationale. Data collection attrition occurs when subjects included in the design are excluded from the data gathering process because they were not contacted by the data gatherer. Data collection attrition also includes subjects who were excluded because the contact with the data gatherer did not yield information required for analysis. Data processing attrition occurs

when subjects included in the design fail to be included in the study data records. Analysis attrition occurs when subjects included in the design are excluded because they are not deemed valid subjects for the analytic design, even though data on them have been correctly collected and are properly included in the study data bank.

In many cases, attrition problems can be dealt with only by documenting and reporting its occurrence to all function groups. There are situations where attrition may be minimized by the way in which each major functional component handles its responsibility for the description and identification of the study subjects.

This paper will be limited to the examination of attrition encountered at the point of data collection. In our discussion, we shall refer to procedures developed during the national longitudinal evaluation of the Follow Through program. These references seem especially relevant since this project was large (over 100 sites) and it was a longitudinal study (covering a period of seven years). The study was organized into design, data collection, data processing, and analysis components.

Major Components of Data Collection Attrition

The data collector gathers information from subjects directly or indirectly by supervising subjects in the completion of data instruments. The data collector must identify the subjects, obtain their cooperation, and perform and document the data collection task in a manner that is consonant with the theoretical design. The data collector must also see that the information gathered is turned over to the data processor in usable form.

For the purposes of this discussion, we shall divide attrition at the point of data gathering into two areas, subject loss, and data loss attrition.

Subject Loss

Subject loss shall be defined as attrition due to the inability of the data collector to find the subject, or inability to obtain permission to work with the subject.

There are some subject losses that the data collector cannot control. They include loss by absence, or movement of the subject beyond the physical reach of the data collector.

Subject attrition may be minimized in the field by gaining the cooperation of subjects, by systematic subject identification and tracking and by field management.

Subject Cooperation

One of the first things a data collection staff must do in the field is to obtain the cooperation of the subjects and those persons who control the environment in which the data are collected. Cooperation is best gained by making sure that all persons involved understand what the research is designed to accomplish and how, why, and by whom data collection will be done. Furthermore, both subjects and relevant persons should know what is expected of them, and what benefits they may gain from working with the study.

Failure to provide correct information about study design and procedures may cause distrust of the data collector--distrust that could easily have been dispelled if the lines of communication had been open. The Follow Through evaluation involved a wide range of stakeholders: parents, children, schools, administrators, practitioners, and program implementation sponsors. The evaluation involved not only the Follow Through community but also persons from the comparison school communities. The task of properly informing such a large and diverse group included several kinds of communication at different levels.

First, we published a letter well before each data collection period. It was mailed to all official stakeholders at all levels (state, local, program) and was made available to anyone who wished a copy. We included not only theoretical material, but data collection schedules, descriptions of instruments, and so on. This was followed up with telephone calls and personal visits to the site by our data collection staff.

In studies involving many cities, the data collector may be the only direct contact subjects have with the research project. Field staff personnel must not only be skilled in working with people, they must also understand the study plans and procedures. They should be able to distinguish those questions best answered by others (analysts, designers, etc.).

The subjects on whom most of the data were collected were children in the primary grades. The mode of data collection used most often was testing. We were very much aware of the apprehension of teachers, parents, and children alike about the testing process. We recognized our obligation to maintain standard testing procedure, but we did not wish to lose subjects because of their fear or dislike of the testing situation. Since our subjects were quite young (kindergarten through third grade), we took great steps to provide a nonthreatening testing situation. Testers spent a significant amount of time on rapport. We explained to classroom teachers why the testing was done and generally how it was being used in the evaluation. With the approval of the teacher, the testers spent time talking and playing informally with the children in their classroom.

By the first day of testing, the SRI staff and the class were acquainted with c another. We hired bilingual testers where necessary and encouraged them to speak to the children in their native language during rapport and pretest practice sessions. Testers were required to learn the children's names and to talk with them about their activities and interests before the beginning of testing. We set up testing guidelines that allowed testers to encourage the children to answer test questions. We also insisted that the children be given as much time as test directions permitted to answer questions.

Testing usually took approximately one hour a day over a three- to five-day period. This long period of time increased the probability of a child's missing a testing session. However, our use of test teams gave us increased ability to deal with make-up testing.

Feedback of data or tentative study results can be very helpful in maintaining good relationships and obtaining cooperation. In a longitudinal study where results extend over a period of years, early data and feedback may be of questionable scientific value. In the early days of the Follow Through evaluation, feedback on data collection was not available. We noted a great improvement in our rapport with local stakeholders once reports and data were available.

We found that providing other kinds of benefits was also helpful. An honorarium was offered to all teachers and aides who were involved in our survey activities. We began offering an honorarium to comparison schools based on the number of subjects tested.

Subject Identification

The data collection staff must locate and document the identity of subjects in the field. In Follow Through, the identification of subjects in the field was done by means of a document called a "roster." A set of rosters was completed for each class involved in the evaluation. On this instrument, name, birthdate, and sex of each subject were recorded. A copy of this form without subject names was returned to SRI. The copies that included subject names were left on site.

In longitudinal studies like Follow Through, the subject identification process must be repeated by each major functional component, at each data collection point. The identification system must be designed to allow the identification of subjects in at least two important ways, the location and identification of the subject at each data collection point, and in terms that will allow each data point to be linked to form a set of data points for that subject.

The Follow Through evaluation data collection and data processing staff developed a system that enabled such links to be made. Each subject was assigned a code at baseline data collection that was used throughout the study. In addition, at each field identification of the subject, both codes were added to the roster form. These were used in data processing to facilitate data links.

Tracking

After baseline data collection, longitudinal samples are usually defined not by a class of subjects, but by subjects who represent a subset of those involved in the baseline data collection. This happens as a result of changes in definition of study subjects by designers and analysts, based on changes in the study environment and goals.

In Follow Through baseline data collection, classes were selected for the sample. That is, the design staff specified kinds of classrooms (i.e., program and comparison) for the sample. The data collection staff collected data on all children in those classrooms.

In subsequent years, additional criteria were defined by the design staff to identify subjects:

(1) Subjects tested in comparison classes at the time of baseline testing must have remained in comparison classes.

This criterion was added because some children were in the Follow Through program for a period of time and then moved to non program classes. These moves were often caused by desegregation, changes in school boundaries, and the movement of families. Also, as the Follow Through program grew, some comparison and non program classes were converted to Follow Through classes. This "crossover" effect caused major losses is some communities.

(2) Subjects must not have been retained at a lower grade or allowed to skip a grade.

The addition of these two criteria eliminated some subjects in the original sample.

Prior to each subsequent data collection, the data collection staff attempted to find and document the location of every member of the original classes. In this way, subjects were eliminated on the basis of the criteria defined by the study designer, as documented in the field by the data collection staff. This process not only minimized attrition, but it discouraged collection of data on ineligible subjects.

Field Management

Once study subjects have been located and identified, arrangements must be made for data collection. When subject are scattered geographically, or are constantly moving about, logistics become very important. As a result of tracking, we learned that:

- The rate of loss of subjects varied among sites and among schools in sites, and that
- In general, program (Follow Through) subjects scattered considerably less than comparison subjects.

For example, in one instance, baseline data collection had occurred in four comparison classes in three schools. Four years later, these subjects were in twenty classrooms in twelve schools.

Consequently, we elected to gather data on the entire program population and on as many of the nonprogram subjects as possible.

To make sure that selected subjects were included in data gathering, the data processing staff provided the field operations staff with a list of target children scheduled for data collection. The field staff checked this list on site for errors and omissions.

Data Loss

Data loss shall be defined as the inability of the data collectors to obtain data necessary to include a subject in data processing and/or analyses plans. Data loss may be minimized by quality control mechanisms, to ensure that each instrument is correctly and fully completed, and by keeping to a careful schedule that makes maximum use of the allotted data collection time.

Data Quality Control

Ideally, every item on every data collection instrument should be completed. An explanation should be given for every item not completed. When several instruments are used, careful planning is necessary to minimize data loss.

Our Follow Through system included such procedures as packaging materials in the order in which they were used, color coding of booklets, administrative logs, and worksheets to document the administration of instruments and the reasons for missing data. We also placed labels on each instrument showing data processing subject identification codes. In addition, data were reviewed by the data collector and data collection supervisors for errors and omissions. We also encouraged children to complete as many items as possible during testing.

Data Collection Scheduling

A good data collection schedule helps to minimize data loss. A good schedule should:

- Provide a reasonable length of time for data collection.
- Take into account environmental situations that may affect data collection.
- Provide time in addition to the regular schedule for collecting data on subjects or records that were unavailable during the regularly scheduled time.

Conclusions

Although attrition is a fact of life in longitudinal studies, the care with which subjects are identified and documented may help to minimize some of its effects.

Subject or data losses in one area may affect other project functions. For this reason, it is important that cases of attrition be communicated to all functional components.

The data collection staff can play an important role in minimizing attrition in longitudinal studies by gaining the cooperation of subjects, by correctly locating and identifying subjects, and by the use of proper management and administration procedures. The careful performance of these tasks lowers attrition rates not only at the point of data collection, but in other functional areas as well.